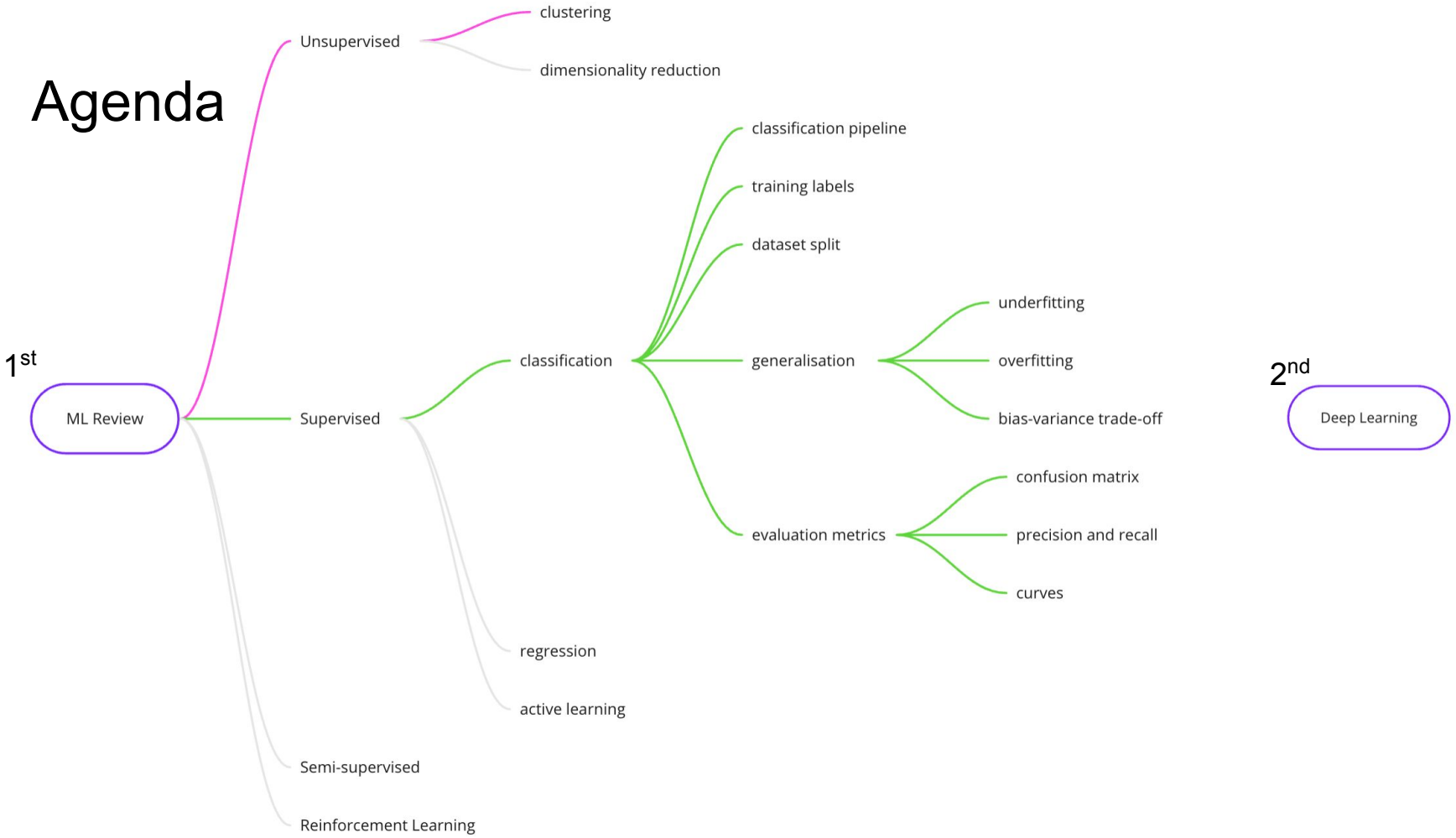
Machine Learning

Machine Learning Review

Agenda


1st
ML Review
Unsupervised
clustering
dimensionality reduction
Supervised
classification
classification pipeline
training labels
dataset split
generalisation
underfitting
overfitting
bias-variance trade-off
evaluation metrics
confusion matrix
precision and recall
curves
regression
active learning
Semi-supervised
Reinforcement Learning
2nd
Deep Learning

Can you provide examples of each one of the following areas?

Supervised Learning

- Classification
- Regression
- Active Learning
- Time-series forecasting
- …

Unsupervised Learning

- Clustering
- Dimensionality reduction
- Anomaly Detection
- Generative models
- …

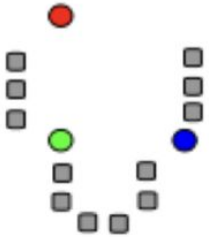
Semi-supervised Learning

- Generative models
- When labels are limited (includes use cases from supervised and unsupervised learning)

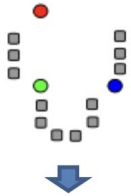
Reinforcement Learning

- Game playing
- Robotics
- Recommendation systems
- …

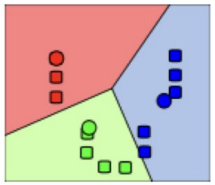
One simple clustering algorithms is k-means, can you describe it?



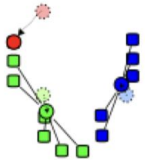
1. Randomly select K centers



2. Assign each point to nearest center

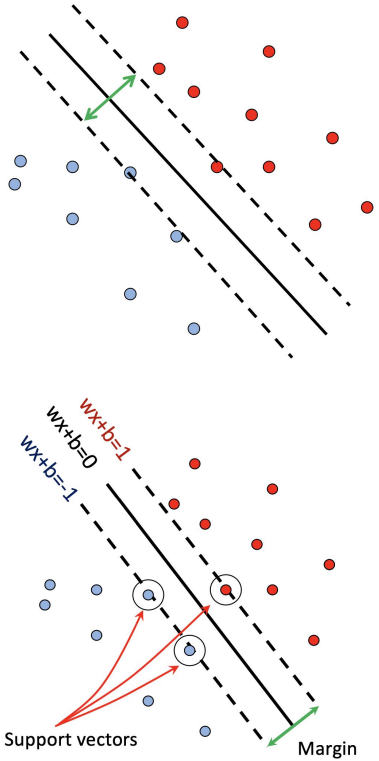


3. Compute new center (mean) for each cluster

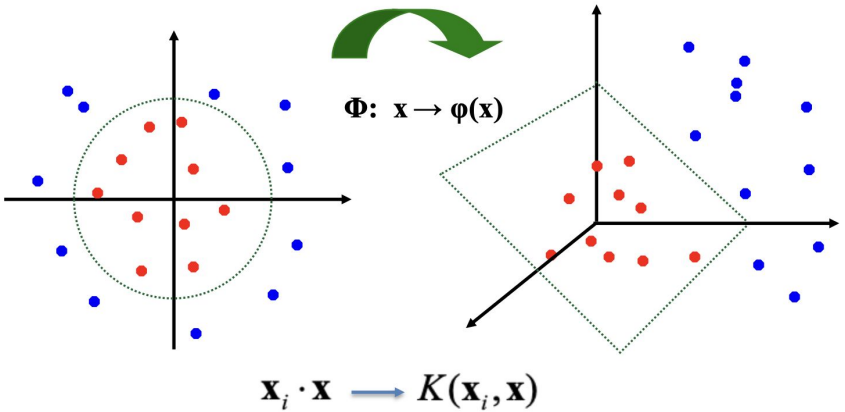


Back to 2

One simple classification algorithm is SVM, can you describe it?



What if the data is not linearly separable?



$$\mathbf{x}_i \cdot \mathbf{x} \longrightarrow K(\mathbf{x}_i, \mathbf{x})$$

Kernel trick

What if the problem is a multiclass problem?

One vs. all

- Training: learn an SVM for each class vs. the rest
- Testing: apply each SVM to test example and assign to it the class of the SVM that returns the highest decision value

One vs. one

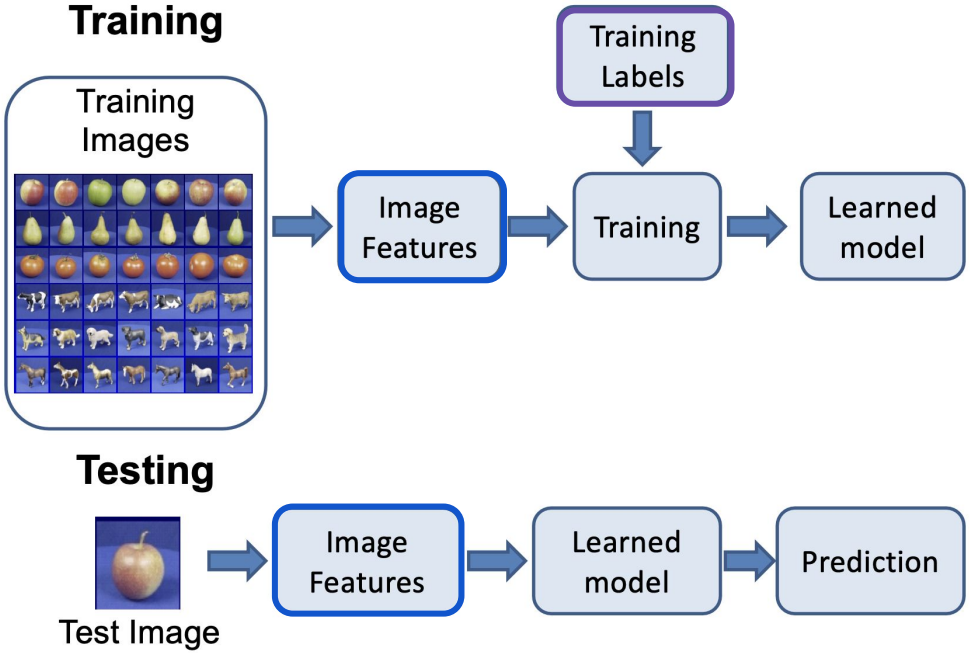
- Training: learn an SVM for each pair of classes
- Testing: each learned SVM “votes” for a class to assign to the test example

What classifiers are you familiar with?

- K-nearest neighbour
- Naïve Bayes
- Bayesian network
- Logistic regression
- Random Forests
- Boosted Decision Trees
- Restricted Boltzmann Machines
- Neural Networks
- Convolutional Neural Networks
- Transformers
- …

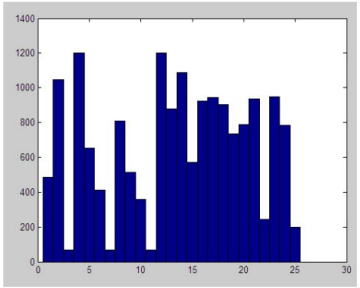
Which one is the best?

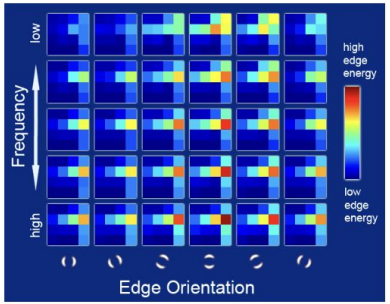
Designing an experiment



What are examples of image features?

- Raw pixels
- Histograms
- Local descriptors
  - SIFT
  - ORB
  - SURF
  - HOG…

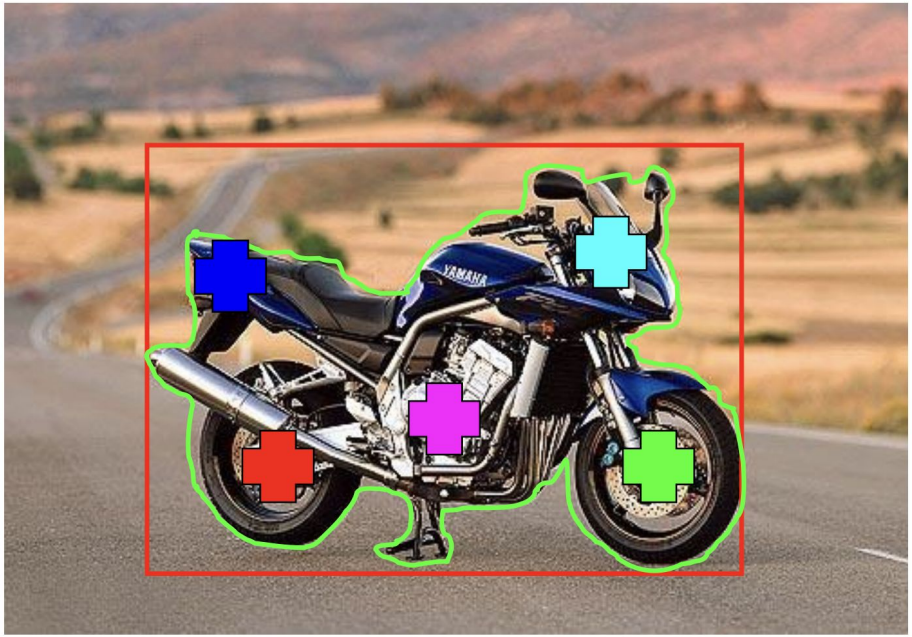




What are examples of training labels?

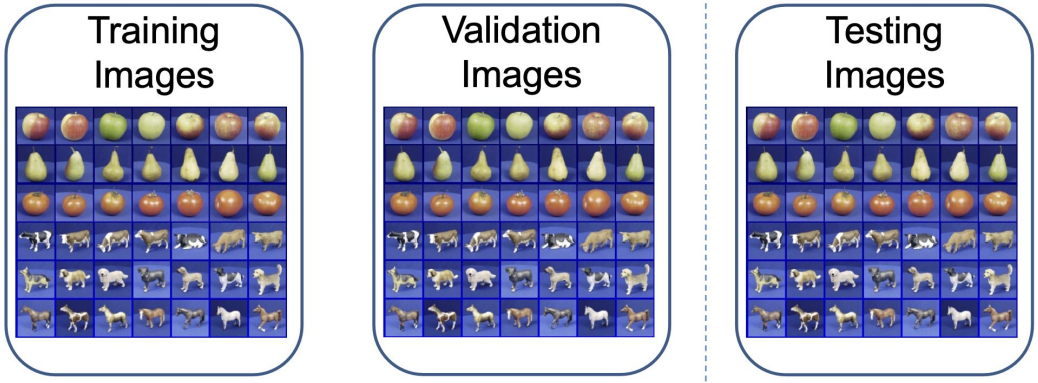
Images in the training set must be annotated with the “correct answer” that the model is expected to produce.

Contains a motorbike



Lana Lazebnik

Assuming we have the data…
What is one of the first tasks to ensure model generalisation?

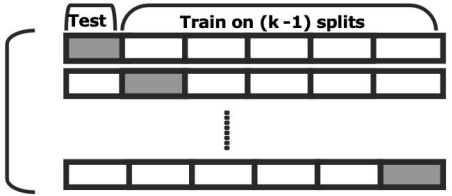


- Train classifier

- Measure error
- Tune model hyperparameters

- Secret labels
- Measure error

Random training/validation splits = cross validation



Generalisation

How well does the trained model generalise to unseen data (test data)?

Training set (labels known)

Test set (labels unknown)

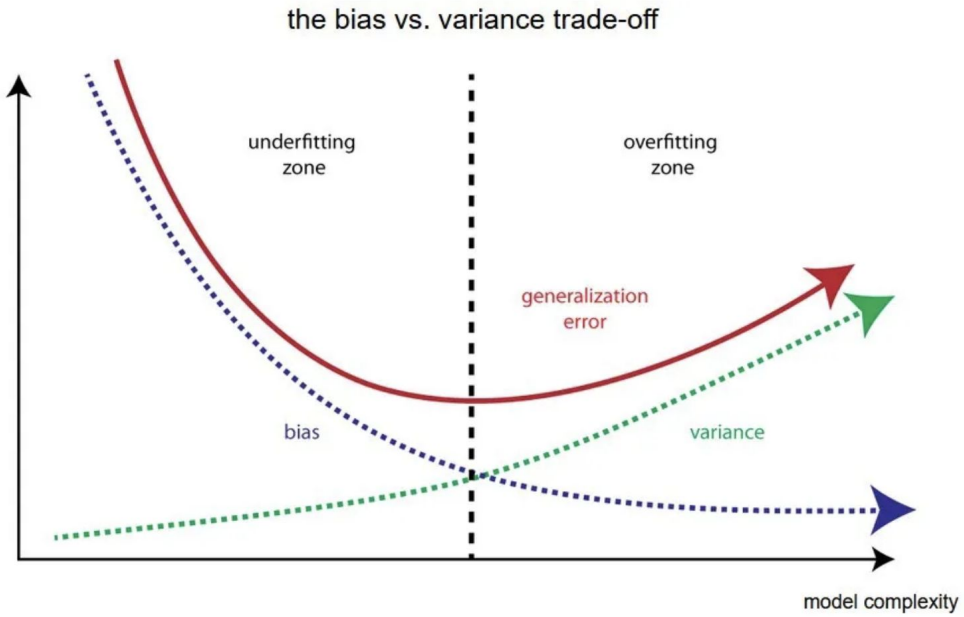
Generalisation error

Bias

- Difference between the expected prediction of our model and the correct value
- Error due to inaccurate assumptions/simplifications

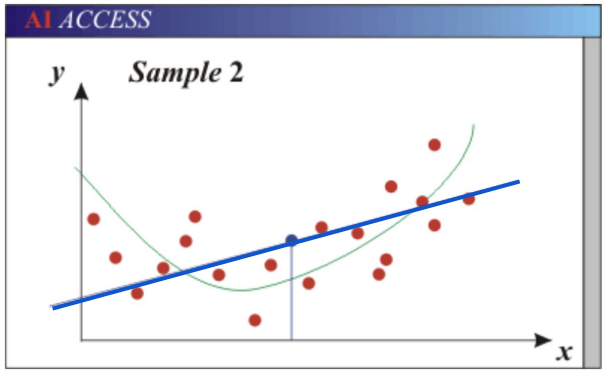
Variance

- How much the target estimates change if different training data is used



Source: https://medium.com/@rsehrawat75/bias-variance-tradeoff-f0e3afb78879

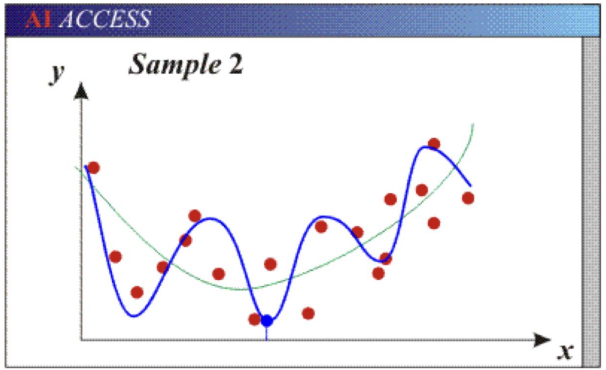
What effects can the generalisation error create?



Underfitting (in blue)

Model is too ‘simple’ to represent all the relevant class characteristics

- High bias (few degrees of freedom) and low variance
- High training error and high test error

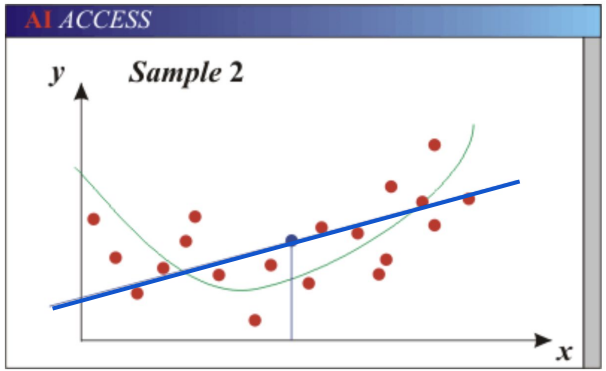


Overfitting (in blue)

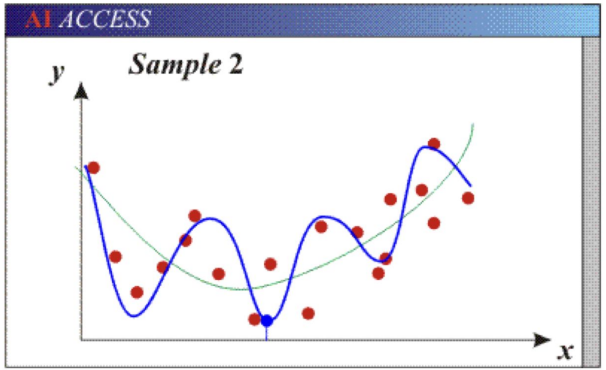
Model is too ‘complex’ and fits irrelevant characteristics (noise) in the data

- Low bias (many degrees of freedom) and high variance
- Low training error and high test error

Bias variance trade-off



- Not enough flexibility
- Too many assumptions



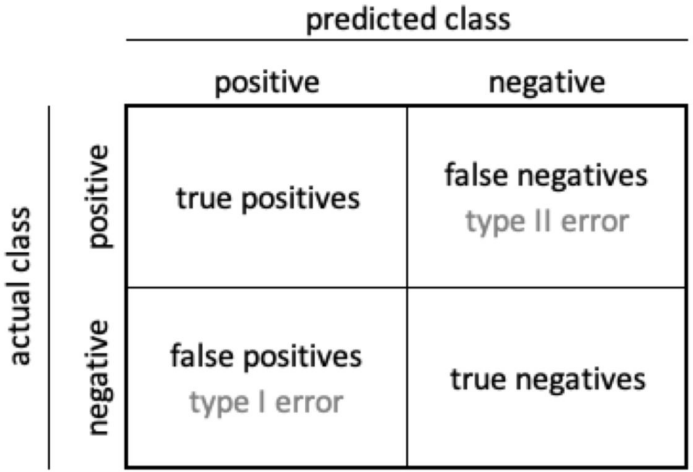
- Too much sensibility to the sample data
- Slightly different data – very different function

Derek Hoiem

How can we evaluate the performance of a classification model?

Confusion Matrix

- For a binary classifier



How can we evaluate the performance of a classification model?

Precision and **Recall**

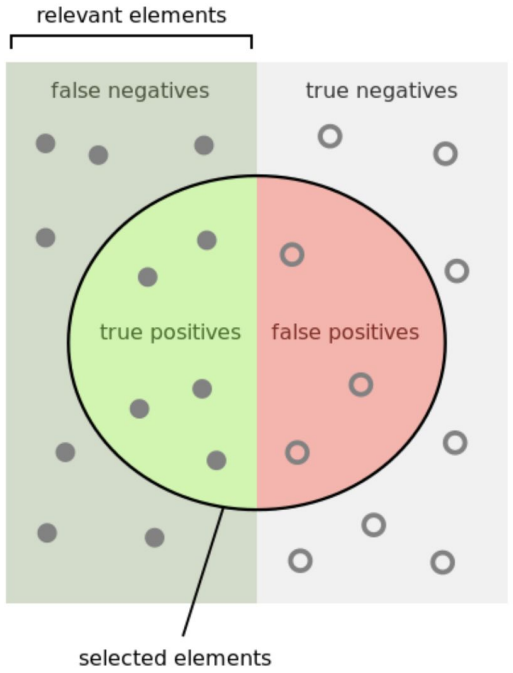
- TP (true positive) - correct attribution
- TN (true negative) - correct rejection
- FP (false positive) - incorrect attribution
- FN (false negative) - incorrect rejection

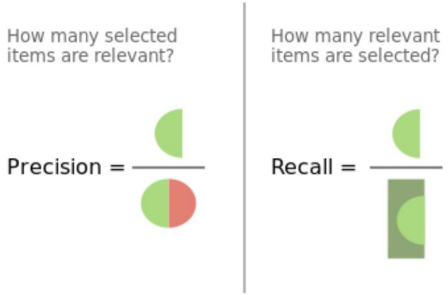
$$\text{Precision} = \frac{tp}{tp + fp}$$

Precision = #relevant / #returned

$$\text{Recall} = \frac{tp}{tp + fn}$$

Recall = #relevant / #total relevant





Curves

Plotting metric values for different threshold values

Receiver operating characteristics (ROC)

- Sensitivity (TPR (true positive rate) or Recall)
- Specificity (TNR (true negative rate))

Area Under the Curve (AUC)

- Also used as a metric

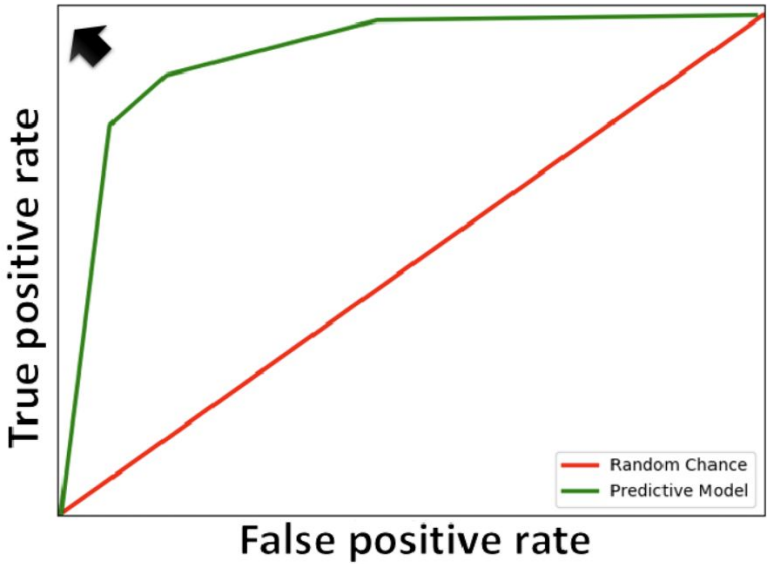


Image source:
https://www.researchgate.net/figure/Representation-of-a-ROC-curve-Ideal-model-marked-by-an-arrow-hypothetical-curve_fig5_324923635

https://xkcd.com/1838/